Notationen:  
 x Inputdaten -> D

z latente Repräsentation , latente Variable

𝜇 Mittelwert latente Verteilung

sigma Standardabweichung latente Verteilung

sigma^2 Varianz

𝑝 (z) prior Verteilung der latenten Variablen( 𝑝𝜃 (𝑥) = N (x; 0, 𝐼) )

𝑞\_𝜙 (𝑧, 𝑥) variable Verteilung

𝑞\_𝜙 (𝑧, 𝑥) = N (𝑧; 𝜇(𝑥), 𝜎2 (𝑥)) Normalverteilung

𝑞\_𝜙 (𝑧|𝑥) approximierte posterior Gauß Verteilung, durch Encoder Modell parametrisiert

q\_ϕ​(z,x)=q\_ϕ​(z∣x) ⋅ q\_ϕ​(x)

𝑝\_𝜃 (𝑥 |𝑧) likelihood-Verteilung der Daten gegeben die latenten Daten (durch decoder Parametrisiert)

𝑝(𝑧|𝑥) wahre posterior VErteilung

𝑞\_𝜙 (𝑥) einfachere Verteilung die wahre posterior approx.

X zufallsvariable

p(x) Wahrscheinlichkeitsdichte

𝐷𝐾𝐿 (𝑞𝜙 (𝑧|𝑥)||𝑝(𝑧)) Kullback-Leibler Divergenz

p(z) prior gauß verteilung

log 𝑝\_𝜃 (𝑥) log likelihood der daten

L\_𝜃,𝜙 (𝑥 ) ELBO

𝜃 Parameter Decoder

𝜙 Parameter Encoder

E\_(q\_ϕ​(z∣x)​) [logp\_θ​(x∣z)] Erwarteter log-Likelihood Term (Rekonstruktionsloss)

Chi Quadrat Test für Herzgesundheit mit 5 EFF: 0.030855528351692245

Abstrakt:

This thesis examines the potential for bias in inference when utilising mixed models in latent representations. The research primarily focuses on the application of Variational Autoencoders (VAE) combined with mixed models to analyse high-dimensional medical datasets. The study investigates the extent to which inference results may be biased and evaluates the robustness of the models used. The objective of this study is to examine the likelihood-ratio test statistics and compare them across different model configurations in order to gain insights into the reliability of using VAEs in conjunction with mixed models for statistical analysis.

Katie Verbesserung:

Quellen in Einleitung

Dopplung Einleitung und kapitel 2

VAE Abkürzung

Seitenzahlen links und rechts